Evaluating Implementation of the Emergency Severity Index in a Belgian Hospital

Authors: Jochen Bergs, MSc, RN, Sandra Verelst, PhD, MD, Jean-Bernard Gillet, MD, and Dominique Vandijck, PhD, RN, Hasselt, Leuven, and Ghent, Belgium

Introduction: Triage aims to categorize patients based on their clinical need and the available departmental resources. To accomplish this goal, one needs to ensure that the implemented triage system is reliable and that staff use it correctly. Therefore this study assessed the ability of Belgium nurses to apply the Emergency Severity Index (ESI), version 4, to hypothetical case scenarios after an educational intervention.

Methods: An ESI educational intervention was implemented in accordance with the ESI manual. Using paper case scenarios, nurses' interrater agreement was assessed by comparing triage nurse ESI levels with the reference answers noted in the implementation manual. Interrater agreement was measured by the percentage of agreement and Cohen's κ coefficient using different weighting schemes.

Results: Overall, 77.5% of the scenario cases were coded according the ESI guidelines, resulting in a good interrater agreement (κ = 0.72, linear weighted κ = 0.84, quadratic weighted κ = 0.92, and triage-weighted scheme = 0.79). Interrater agreement varied when evaluating each ESI level separately. Undertriage was more common than overtriage. The highest misclassification range (37.8%) occurred in ESI level 2 scenarios, with 99.2% of the misclassifications being undertriaged.

Discussion: Implementation of the ESI into a novel setting guided by a locally developed training program resulted in suboptimal interrater agreement. Existing weighted κ schemes overestimated the interrater agreement between the triage nurse–assigned ESI level and the reference standard. By providing an aggregated measure of agreement, which allows partial agreement, clinically significant misclassification was masked by a misleading "good" interrater agreement.

Key words: Emergency nursing; Emergency service, hospital; Triage; Implementation

Emergency departments are increasingly confronted with situations in which their function is impeded when the number of patients who are waiting to be seen, undergoing assessment and treatment, or waiting for discharge exceed the physical and/or staffing capacity of the emergency department.[1,2] The problem does not end at the ED door; hospitals in general are getting saturated, resulting in unreasonable waiting times before ED patients can be transferred to a staffed hospital bed.[1-3] These situations, known as crowding and access block, cause extended ED waiting times that potentially jeopardize patient safety.[4-6]

A frequently used method to prevent unsafe waiting times is to determine clinical priorities among visiting patients. Several urgency classification methods are available within the literature.[7] One example is the Emergency Severity Index (ESI), a 5-level triage scale.[8,9] Essentially, triage is a process of sorting patients into meaningful groups. These groups can be used to manage the waiting patients by giving priority to certain groups or streaming patients according their needs. The overall objective of a triage system is to identify high-risk patients, essentially those who cannot wait to be seen.[7-9] The ESI is based around a new conceptual model of ED triage. It retains the

Jochen Bergs is PhD candidate, Patient Safety Research Group, Faculty of Business Economics, Hasselt University, Hasselt, Belgium.

Sandra Verelst is Emergency Physician, Department of Emergency Medicine, Leuven University Hospitals, Leuven, Belgium.

Jean-Bernard Gillet is Professor, Department of Emergency Medicine, Leuven University Hospitals, Leuven, Belgium.

Dominique Vandijck is Professor, Patient Safety Research Group, Faculty of Business Economics, Hasselt University, Hasselt, Belgium, and Professor, Department of Public Health and Health Economics, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium.

For correspondence, write: Jochen Bergs, MSc, RN, Martelarenlaan 42, 3500 Hasselt, Belgium; E-mail: kate.moore@emory.edujochen.bergs@uhasselt.be.

J Emerg Nurs 2014;40:592–7.
Available online 12 March 2014

0099-1767

Copyright © 2014 Emergency Nurses Association. Published by Elsevier Inc. All rights reserved.

http://dx.doi.org/10.1016/j.jen.2014.01.006

traditional foundation of patient urgency (ie, sorting of patients) while seeking to accomplish a second goal of patient streaming: that is, getting the right patient to the right resources at the right place and time.[10] The ESI attempts to accomplish these 2 goals by indicating which patients should be seen first and, additionally, considering the resources required to determine the patient's disposition.[9,10] Triage is intended to ensure patient safety among waiting patients, and thus accurate triage is fundamental, especially because the assigned triage level determines the waiting time and initial level of care a patient will receive.[11] Therefore the triage system needs to be valid and reliable at the same time. Several studies indicate sufficient to excellent validity and reliability of the ESI.[12-20]

Implementation of the ESI has primarily taken place in the United States, although some European countries have adopted this system as well.[21] In Belgium, nurse triage has only been legally possible since 2007. Subsequently, many Belgian hospitals have been persuaded to implement a triage system. For the previously stated reasons, emergency departments are inclined to implement a triage system with demonstrated validity and reliability. After implementation of a new system, formal evaluation of its performance within the new setting is recommended.

The objectives of this study are 2-fold. First, we evaluated the ability of Belgian nurses to apply the ESI to hypothetical paper-based case scenarios according to an educational program. A second objective of this study was to compare different measures of interrater agreement in hypothetical triage scenarios.

Methods

DESIGN, SETTING, AND PARTICIPANTS

This cross-sectional observational study was conducted at the emergency department of a 1900-bed tertiary care teaching hospital in Belgium. The emergency department has an annual census of approximately 54 000 patients and a census of 102 to 210 patients per day. Between 2005 and 2008, the patient volume increased by approximately 3814 patients (8%). The ESI was implemented in the emergency department in January 2009. The ED staff had no previous experience with the ESI or triage in general. Under Belgian law, only nurses with a license in intensive care and emergency care (gained after obtaining an additional bachelor's degree in intensive and emergency care nursing) are allowed to perform triage in the emergency department. During the period from December 2008 to March 2009, all emergency nurses allowed to perform triage were trained (N = 52). Besides having a license in intensive care and emergency care, all participating nurses had at least 2 years of work experience in the emergency department.

PROCEDURE

A multidisciplinary implementation team consisting of 3 emergency physicians and 4 emergency nurses developed and provided an education program to teach nurses to use the ESI triage standards (version 4). The content was obtained from the ESI implementation manual.[10] Educational support came from one of the team members, a qualified teacher in Medico-Social Sciences. The resulting training program consisted of a 3-hour interactive theoretical presentation, followed by a practice session with 30 paper-based case scenarios, also adopted from the implementation manual (chapter 9: practice cases).[10] The original implementation manual was not distributed among the nurses. Instead, handouts of the theoretical presentation and a reference card were provided. To succeed in the training program, all participating nurses had to complete a survey consisting of 30 paper-based case scenarios, which were also adapted from the implementation manual (chapter 10: competency cases).[10] Prior to implementation of the training program, 2 independent medical management assistants translated the practice and competency cases into Dutch. The translated cases were translated back to English by the implementation team to correct translation inconsistencies. All team members had excellent knowledge of English.

The training program was spread over 2 days, dividing theoretical and practice sessions, allowing the nurses to process the material. The survey was conducted after the practice session. Nurses were allowed to use the reference card containing the ESI algorithm and a summary of the resources as defined in the ESI algorithm, version 4.

DATA ANALYSIS

The survey answers (triage nurse–assigned ESI level) were compared with the reference answers provided in the implementation manual (true ESI level).[10] We defined undertriage as the assignment of a lower triage level compared with the reference answer. Overtriage was defined as assignment of a higher triage level compared with the reference answer. Based on a contingency table, data were descriptively analyzed with use of frequencies and percentages. Because of the lack of international consensus regarding the evaluation of interrater agreement in triage, several statistical measures were used. Cohen's κ coefficient is a frequently used measure; it is a statistical measure in which agreement of 2 or more raters or methods (interrater

agreement) is corrected for agreement expected by chance. The chance correction in κ depends on the distribution of ratings and the number of categories of the measurement scale; κ ranges between −1, representing perfect disagreement, and 1, representing perfect agreement.[22,23] The unweighted scheme is used for measuring exact agreement in nominal scales. Weighted schemes also measure partial agreement and are used for ordinal scales. Weights can be calculated with a linear or quadratic algorithm or can be defined by the user.[22,23] Because triage systems are ordinal scales, weighted κ seems to be the correct measure for evaluating interrater agreement. Some authors suggested that weights used for the calculation of weighted κ in triage interrater agreement studies should take the severity of misclassification into account.[24,25]

In this study, interrater agreement was assessed using percentage of agreement, linear-weighted kappa (κ_{lw}), quadratic-weighted kappa (κ_{qw}) and triage-weighted scheme (κ_{triage}) as reported by van der Wulp and van Stel.[24] Interpretation of the obtained κ values was based on the definitions reported by Altman[26] and Viera and Garrett.[27] For data management and analysis, R, version 2.13.1 (R Development Core Team, Vienna, Austria) and IBM PASW Statistics 18 (SPSS Inc, Chicago, IL) were used.

ETHICAL CONSIDERATIONS

Prior to participation, all participants were verbally informed about the objectives of the survey. Consent was assumed if participants completed the survey.

Results

The mean score for the survey was 23/30. Overall exact agreement between triage nurses and reference answers was 77.5%. The mistakes made by triage nurses resulted more commonly in undertriage (77.5%) than overtriage. Undertriage ranged from a difference of 1, 2, or 3 levels from the true ESI level; overtriage ranged from a difference of 1 or 2 levels from the true ESI level. The unweighted κ score was 0.72 (95% confidence interval [CI] 0.69-0.74), κ_{lw} was 0.84 (95% CI 0.77-0.91), κ_{qw} was 0.92 (95% CI 0.86-0.98), and κ_{triage} was 0.79 (95% CI 0.69-0.89; Table 1).

Table 2 shows the results for each ESI level separately. Errors were made in all ESI levels, with the highest error range in ESI level 2. ESI level 1 scenarios were correctly rated in 87.5% of cases. Incorrect answers within this triage level were all rated as ESI level 2. ESI level 2 scenarios were correctly rated in 62.2% of cases. Incorrect answers were mainly undertriaged (99.2%), and varied between ESI level 3 (83.76%), ESI level 4 (13.67%), and ESI level 5 (2.56%).

TABLE 1
Comparison of triage nurse-assigned ESI levels to true ESI levels

True ESI level	Triage nurse-assigned ESI level					Total No. of scenarios
	ESI 1	ESI 2	ESI 3	ESI 4	ESI 5	
ESI 1	273	39	0	0	0	312
ESI 2	1	194	98	16	3	312
ESI 3	1	45	265	40	13	364
ESI 4	0	3	6	240	63	312
ESI 5	0	0	2	21	237	260
Global	275	281	371	317	316	1560

ESI, Emergency Severity Index.

Overtriage occurred in only one case. ESI level 3 scenarios were correctly rated in 72.8% of cases. Incorrect answers were mainly undertriaged (53.5%), and varied between ESI level 4 (75.5%) and ESI level 5 (24.5%). Overtriage occurred in 46.5% of cases, and varied between ESI level 1 (2.17%) and ESI level 2 (97.82%). ESI level 4 scenarios were correctly rated in 76.9% of cases, with incorrect answers mainly being undertriaged (87.5%). Overtriage occurred in 12.5% of cases and varied between ESI level 2 (33.33%) and ESI level 3 (66.67%). Finally, ESI level 5 scenarios were correctly rated in 91.1% of cases. Incorrect answers varied between ESI level 3 (8.7%), and ESI level 4 (91.3%).

Discussion

Triage is a vital tool in preventing unsafe waiting times for the most vulnerable patients. However, to accomplish this goal, the implemented triage system needs to be valid and reliable and staff need to use the tool correctly. Because triage relies on a nurse's interpretation of the level specific criteria, interrater variability is problematic but surmountable. In other words, the interrater agreement of a triage system is dependent upon nurses' knowledge and experience, as well as their understanding of and ability to use the triage algorithm correctly. This situation implies that all aspects of implementation of a triage system should be formally evaluated, including the education of nurses who will use the system. To that end, this study assessed the interrater agreement of nurses' use of the ESI version 4 in a single tertiary care teaching hospital in Belgium by using ED triage scenarios.

Linear and quadratic weighted κ schemes, which provide an overall score of interrater agreement, suggest a

TABLE 2
Individual Emergency Severity Index level analysis

True ESI level	No. correct (%) [a]	No. overtriaged (%) [b]	No. undertriaged (%) [b]
ESI 1	273 (87.5)	–	39 (100)
ESI 2	194 (62.2)	1 (0.8)	117 (99.2)
ESI 3	265 (72.8)	46 (46.5)	53 (53.5)
ESI 4	240 (76.9)	9 (12.5)	63 (87.5)
ESI 5	237 (91.1)	23 (100)	–
Overall	1209 (77.5)	79 (22.5)	272 (77.5)

ESI, Emergency Severity Index.
[a] Percent of total scenarios.
[b] Percent of misclassified scenarios.

"very good" agreement between triage nurse-assigned ESI level and the reference standard (κ_{lw} = 0.84 and κ_{qw} = 0.92). Triage-weighted kappa suggests "good" agreement (κ_{triage} = .0.79). These values are in line with previously reported data.[9,12,14,15,17-20] However, consideration of the results for each ESI level separately has some important implications. The results of this study suggest that nurses had difficulties with the interpretation of high acuity level criteria and estimating the number of resources a patient needs to reach a dispositional decision. The criteria to assign a patient to ESI level 1 are clearly defined and therefore more tangible than lower acuity levels, which is reflected by adequate agreement between triage nurses' ESI level and the true ESI level (87.5%). In contrast, the criteria for ESI level 2 are less specifically formulated and rely heavily on the knowledge and experience of ED nurses; as a result, poor agreement was found between triage nurse–assigned ESI level and the true ESI level (62.2%). Disagreement, except for one case, resulted in undertriage, mainly as ESI level 3. Because ESI level 2 patients remain a high priority, assigning these lower acuity levels could jeopardize patient safety. The distinction between ESI levels 3, 4, and 5 is based on the estimated number of resources needed to determine the patient's disposition. This estimation is based on the "gold standard" of care a patient would normally receive given his or her condition. To identify ED patients' resource needs, the triage nurse must be familiar with general ED standards of care, and specifically with what constitutes prudent and customary emergency care.[10] In this study, more errors were made in ESI level 3 scenarios (27.2%) compared with ESI level 4 scenarios (23.1%) and ESI level 5 scenarios (8.8%).

As demonstrated in this study, poor results in one ESI level can be masked by the results of other levels, resulting in an overall "good" interrater agreement. Bearing in mind that the assigned ESI level determines waiting time and the initial level of care a patient will receive, we need a measure of interrater agreement that reflects the severity of misclassification.[24] If we compare the obtained κ values and interpret them in relation to the performance of each individual ESI level, we believe that the κ values insufficiently express the severity of misclassification. Weighted κ is suggested as the most appropriate κ given the ordinal nature of triage instruments. However, weighted κ allows partial agreement. Therefore van der Wulp et al.[24] suggested that the weights must fit to what is clinically more acceptable in triage practice. Their weighting scheme is based on two principles. First, undertriage, especially of ESI level 1 and 2 patients, can cause unnecessary morbidity and mortality. Second, in cases of overtriage, the waste of ED resources must be prevented because it can also cause prolonged waiting times for truly urgent patients. Therefore several cells were set to zero. In the least urgent triage category, 2 levels of overtriage were allowed because overtriage of the least urgent patients is not likely to have a major impact on the speed of treatment onset for patients in the urgent triage categories. By using triage-weighted κ, van der Wulp et al[24] believe that interrater agreement of triage systems will be interpreted in accordance with clinical practice. Further research regarding the adjustment of these weighting schemes to estimate interrater agreement for the ESI system seems necessary so that interrater agreement will be interpreted in accordance with clinical practice and appropriately reflect the severity of misclassification.

These results must be appreciated within the context of the study setting. This emergency department had no prior experience with the ESI. Implementation of the ESI system was based on the publicly available implementation manual, which was used to establish a training program. Our lack of experience with the use of ESI represents the way the vast majority of Belgian hospitals, and probably European hospitals in general, will implement the ESI. The ESI as a tool guides nurses through several decision points and, as such, depends on their knowledge and experience. Although all nurses have an additional degree in emergency nursing and at least 2 years of working experience, most errors made in those ESI levels were dependent upon the knowledge and experience of the triage nurse. When human lives are at stake, a 77.5% correct classification is unacceptable when implementing a new triage system. This finding highlights the importance of education and training during the implementation of a triage system to ensure that the safety of patients is not endangered. Further research is needed to ascertain the amount of training and experience

needed for nurses to be able to use the ESI system accurately and precisely.

Limitations

When interpreting the results of this study, one needs to be aware of several limitations. First, this study was conducted at a single center with a limited number (N = 52) of triage nurses. Second, because the native language of triage nurses is Dutch, the implementation team had to translate the contents of the implementation manual, and thus potential important nuances could have been lost. Third, triage decision making is a process influenced by multiple factors.[12-20] As a result, hypothetical paper-based case scenarios could have led to different results than might have been achieved with real-time patient triage assessments. However, we promote the use of standardized scenarios, as provided in the implementation manual, because it allows the comparison of nurses' understanding of the ESI algorithm. Moreover, it is often not feasible to conduct simultaneous real-time patient triage assessments on single patients with multiple triage nurses.

Implications for Emergency Nurses

The results of this study suggest that Belgian nurses have difficulties with the interpretation of high acuity level criteria, as well as estimating the number of resources a patient needs to reach a dispositional decision. The ESI manual noted that these decisions are primarily based on nurses' knowledge and experience. If a triage system is largely based on knowledge and experience, interobserver variation is unavoidable, which implies that sufficient attention must be directed at the implementation process, particularly the education and training of triage nurses. Misinterpretations of the ESI criteria will affect decisions made by nurses at the front door of the emergency department, making even a valid and reliable triage system ineffective.

Conclusions

In conclusion, implementation of the ESI into a novel setting guided by a locally developed training program resulted in suboptimal interrater agreement of triage classification. The majority (77.5%) of the assigned triage codes by nurses were in concordance with the true ESI level, and errors were made in all ESI levels, with the highest number of errors in ESI level 2. Further, this study shows that existing weighted κ schemes overestimate the interrater agreement between of the triage nurse-assigned ESI level and the reference standard. Poor results for one ESI level were offset by good results in other levels because of the allowance of partial agreement and by providing an aggregated measure of agreement, which resulted in a misleading "good" interrater agreement.

REFERENCES

1. Richardson DB, Mountain D. Myths versus facts in emergency department overcrowding and hospital access block. *Med J Aust.* 2009;190:369-74.
2. Forero R, Hillman K. . Access block and overcrowding: a literature reviewUniversity of New South Wales Simpson Centre for Health Service Research; 2009.
3. Schull MJ, Szalai JP, Schwartz B, Redelmeier DA. Emergency department overcrowding following systematic hospital restructuring: trends at twenty hospitals over ten years. *Acad Emerg Med.* 2001;8:1037-43.
4. Richardson DB. Increase in patient mortality at 10 days associated with emergency department crowding. *Med J Aust.* 2006;184:213-6.
5. Sprivulis PC, Da Silva JA, Jacobs IG, Frazer AR, Jelinek GA. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Med J Aust.* 2006;184:208-12.
6. Cameron PA. Hospital overcrowding: a threat to patient safety? *Med J Aust.* 2006;184:203-4.
7. Mistry RD, Brousseau DC, Alessandrini EA. Urgency classification methods for emergency department visits: do they measure up? *Pediatr Emerg Care.* 2008;24:870-4.
8. Gilboy N, Travers DA, Wuerz RC. Reevaluating triage in the new millennium: a comprehensive look at the need for standardization and quality. *J Emerg Nurs.* 1999;25:468-73.
9. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med.* 2000;7:236-42.
10. Gilboy N, Tanabe P, Travers DA, Rosenau A, Eitel D. Emergency Severity Index, Version 4: Implementation HandbookAgency for Healthcare Research and Quality; 2005. AHRQ, publication 05–0046–2.
11. Ducharme J, Tanabe P, Homel P. The influence of triage systems and triage scores on timeliness of ED analgesic administration. *Am J Emerg Med.* 2008;26:867-73.
12. Baumann MR, Strout TD. Evaluation of the emergency severity index (version 3) triage algorithm in pediatric patients. *Acad Emerg Med.* 2005;12:219-24.
13. Baumann MR, Strout TD. Triage of geriatric patients in the emergency department: validity and survival with the Emergency Severity Index. *Ann Emerg Med.* 2007;49:234-40.
14. Durani Y, Brecher D, Walmsley D, Attia MW, Loiselle JM. The Emergency Severity Index Version 4: reliability in pediatric patients. *Pediatr Emerg Care.* 2009;25:504-7.

15. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The Emergency Severity Index triage algorithm version 2 is reliable and valid. *Acad Emerg Med.* 2003;10:1070-80.
16. Tanabe P, Gimbel R, Yarnold PR, Adams JG. The Emergency Severity Index (version 3) 5-level triage system scores predict ED resource consumption. *J Emerg Nurs.* 2004;30:22-9.
17. Tanabe P, Gimbel R, Yarnold PR, Kyriacou DN, Adams JG. Reliability and validity of scores on the Emergency Severity Index version 3. *Acad Emerg Med.* 2004;11:59-65.
18. Travers DA, Waller AE, Katznelson J, Agans R. Reliability and validity of the Emergency Severity Index for pediatric triage. *Acad Emerg Med.* 2009;16:843-9.
19. Wuerz RC, Travers D, Gilboy N, Eitel DR, Rosenau A, Yazhari R. Implementation and refinement of the emergency severity index. *Acad Emerg Med.* 2001;8:170-6.
20. Worster A, Gilboy N, Fernandes CM. Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. *CJEM.* 2004;6:240-5.
21. Elshove-Bolk J, Mencl F, Van Rijswijck BTF, Simons MP, Van Vugt AB. Validation of the Emergency Severity Index (ESI) in self-referred patients in a European emergency department. *Emerg Med J.* 2007;24:170-4.
22. Cohen JA. Coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37-46.
23. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 1968;70:213-20.
24. van der Wulp I, van Stel HF. Adjusting weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol.* 2009;62:1196-201.
25. van der Wulp I, van Stel HF. Calculating kappas from adjusted data improved the comparability of the reliability of triage systems: a comparative study. *J Clin Epidemiol.* 2010;63:1256-63.
26. Altman D. Practical Statistics for Medical ResearchChapman & Hall; 1991.
27. Viera AJ, Garrett JM. Understanding interobserver agreement: the Kappa statistic. *Fam Med.* 2005;37:360-3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.